Document de Travail
Working Paper

EA 4272

Are Unit Root Tests Useful in the Debate over the (Non)Stationarity of Hours Worked?

Amélie CHARLES (*)
Olivier Darné (**)
Fabien Tripier (**)

2010/27

(*) Audencia Nantes – School of Management
(**) LEMNA - Université de Nantes

Laboratoire d'Economie et de Management Nantes-Atlantique
Université de Nantes
Chemin de la Censive du Tertre – BP 52231
44322 Nantes cedex 3 – France
www.univ-nantes.fr/iemn-iae/recherche
Tél. +33 (0)2 40 14 17 17 – Fax +33 (0)2 40 14 17 49

Are Unit Root Tests Useful in the Debate over the (Non)Stationarity of Hours Worked?

Amélie CHARLES[*], **Olivier DARNÉ**[†], **Fabien TRIPIER**[‡]

October 18, 2010

Abstract

This article compares the performances of some non-stationarity tests on simulated series, using the business-cycle model of Chang et al. (2007) [*Y. Chang, T. Doh, F. Schorfheide, (2007). Non-stationary Hours in a DSGE Model. Journal of Money, Credit and Banking 39, 357-1373*] as data generating process. Overall, Monte Carlo simulations show that the efficient unit root tests of Ng and Perron (2001) [*Ng, S., Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-1554*] are more powerful than the standard non-stationarity tests (ADF and KPSS). More precisely, these efficient tests are able to reject frequently the unit-root hypothesis on simulated series using the best specification of business-cycle model found by Chang et al. (2007), in which hours worked are stationary with adjustment costs.

JEL Classification: C32, E24, E32

Keywords: unit root rest, DSGE models, hours worked

[*]Audencia Nantes, School of Management, 8 route de la Jonelière, 44312 Nantes Cedex 3. Email: acharles@audencia.com.

[†]LEMNA, University of Nantes, IEMN–IAE, Chemin de la Censive du Tertre, BP 52231, 44322 Nantes, France. Email: olivier.darne@univ-nantes.fr.

[‡]Corresponding author: LEMNA, University of Nantes, IEMN–IAE, Chemin de la Censive du Tertre, BP 52231, 44322 Nantes, France. Email: fabien.tripier@univ-nantes.fr.

1 Introduction

Applications of unit root tests, and more generally non-stationarity tests, to financial and macroeconomics series have challenged conventional economic theory and stimulated the development of new theories in numerous fields, such as, economic fluctuations (Nelson and Plosser, 1982).[1] However, a well-known shortcoming of unit root tests restricts this approach. The properties of unit root tests are generally weak for the sample size of typical macroeconomic time series (about 100-200 observations) (Haldrup and Jansson, 2006). The recent debate on the stationarity of hours worked illustrates this shortcoming and brings into question the usefulness of unit root tests in developing business cycle models. In this article, we show how to improve the usefulness of these tests by using economic theory.

The debate over the stationarity of hours worked ensues from the Gali (1999)'s results on the effects of technological shocks[2], which contradict the technology-driven business cycle theory. These results are based on a Structural VAR (SVAR) model *à la* Blanchard and Quah (1989) where the series of hours worked is introduced in first-difference. Gali (1999) motivates this specification by the outcome of standard Augmented Dickey-Fuller (ADF) tests.[3] Among the responses to the Gali (1999)'s findings[4], Christiano et al. (2004) find the opposed results of the author by introducing hours worked in level, and not in first-difference, in the SVAR. As Gali (1999), Christiano et al. (2004) motivate the SVAR specification by the outcome of a stationarity test. Whelan (2009) provides also contradictory results according to the tests and the data

[1] For example, the detection of a unit root in output by Nelson and Plosser (1982) has legitimated the development of business cycle models with highly persistent or non-stationary shocks to factors' productivity. The first generation of Real Business Cycle models has considered a highly persistent autoregressive process for the technological shock – see Kydland and Prescott (1982), Hansen (1985) and Prescott (1986). The process of technological shocks has been modelled as a random walk later, generally in multiple-shocks models as in King et al. (1991) and Christiano and Eichenbaum (1992). See Hansen (1997) for a discussion on this issue.

[2] Gali (1999) concludes that technological shocks play a minor role in the business cycle and that a positive technological shock induces a fall in hours worked.

[3] Gali and Rabanal (2004) extend the set of tests to the KPSS test and confirm the findings of Gali (1999).

[4] For example, Francis and Ramey (2005) develop Real Business Cycle models consistent with a negative response of employment to a positive technological shock whereas Chari et al. (2008) advance that SVAR are useless in developing business cycle theory.

used.

Due to these mixed results, one literature suggests to minor or to abandon the use of standard unit root tests.[5] However, the previous studies on the stationarity of hours worked suffers from two drawbacks. Firstly, they consider few and relatively old tests (namely ADF and KPSS) and do not include the recent developments of efficient unit root tests, especially those of Elliott et al. (1996) and Ng and Perron (2001) (respectively, ERS and NP hereafter). For example, when we apply the ERS and NP tests on the three data sets employed in Chang et al. (2007), we obtain different results on the (non)stationarity of the hours worked than those found by the authors using ADF tests (see Table 1).[6] Secondly, when several tests are used, their relative performances are not assessed in the business-cycle model framework. However, if observed data are viewed as one realization of an economic model, it is essential that the unit root tests used perform well when this economic model is considered as the true data generating process (DGP hereafter).[7] To overcome these limits, we compare herein the performances of several tests (ADF, KPSS, ERS and NP) using a business cycle model as the DGP. More precisely, we choose the model proposed by Chang et al. (2007), which has several key attractive features. It (*i*) allows for either stationary or non-stationary hours worked, (*ii*) does consider or not adjustment costs of labor, and (*iii*) has been estimated with Bayesian methods to account for business cycle facts on output and labor. We use the four model specifications estimated by Chang et al. (2007) to assess the sensitivity of test performances to the choice of the DGP. For each specification, we simulate the model for various sample sizes (100, 200, 500, 1000) and evaluate the size and power properties of the alternative tests.

We show that the performances of the tests are highly sensitive to the specification of the model, i.e.

[5] To overcome the choice between hours worked in level or in first-difference, for example, Fève and Guay (2009) suggest to use a more clearly stationary variable in the SVAR in place of hours, namely the ratio of consumption to output, and show how to recover the responses of hours to shocks in a second step, independently of the specification of the series (in level or in first-difference).

[6] Table 1 illustrates the difficulty in testing for unit root in observed data for hours worked. For two of the three series, the unit root hypothesis is rejected by the ERS and NP tests whereas this hypothesis is not rejected for the three series according to the ADF test.

[7] For example, one issue with standard unit root tests used in Chari et al. (2008) is that they are unable to reject the hypothesis where the hours series has a unit root whereas the hours series in the model is highly persistent, but stationary.

the structure of shocks as well as the existence of adjustment costs. Even if the ADF and NP tests give similar (wrong) properties for the DGP with stationary hours and no adjustment costs of labor, the NP tests strongly dominate the ADF test when the adjustment of labor is costly. This result puts forward the need for of a rigorous assessment of test performances, before applying them to observed data. But, this result also raises the issue of the relevant choice of the model specification given its impact on the evaluation of tests. In the model of Chang et al. (2007), adjustment costs are a powerful propagation mechanism that induce hump-shaped responses of hours worked to shocks. Since adjustment costs are widely supported by quantitative macroeconomic studies, and notably by Chang et al. (2007), among others, these results lead us to prefer the model specification with adjustment costs and therefore to recommend the NP tests instead of the ADF test.

The remainder is as follows. Sections 2 and 3 describe the (non-)stationarity tests and the business-cycle models as DGPs, respectively. Section 4 presents the Monte Carlo simulation design and the results. Section 5 concludes.

2 The Unit Root and Stationarity Tests

We present the tests used in this study and provide an application to observed data. We assume that the data $y_1, \ldots, y_T$ were generated as

$$\begin{aligned} y_t &= \psi' z_t + u_t \qquad\qquad (t = 1, \ldots, T), \\ u_t &= \alpha u_{t-1} + \upsilon_t \end{aligned}$$

where z_t is a deterministic component and υ_t is an unobserved stationary zero-mean error process.

2.1 Elliott, Rothenberg and Stock (1996)

Elliott et al. (1996) (ERS, thereafter) developed a unit root test based on a quasi-difference detrending of the series in order to increase power of Dickey-Fuller (1979, 1981) tests. They suggest the Dickey-Fuller generalized least squares (DF-GLS) test using the following regression

$$\Delta \tilde{y}_t = \beta_0 \tilde{y}_{t-1} + \sum_{j=1}^{k} \beta_j \Delta \tilde{y}_{t-j} + \varepsilon_t$$

where $\tilde{y}_t$ is the locally detrended series y_t. The DF-GLS t-test is performed by testing the null hypothesis $\beta_0 = 0$ against the alternative $\beta_0 < 0$. The local detrending series is defined by

$$\tilde{y}_t = y_t - \hat{\psi}' z_t$$

where z_t equals to 1 for the constant mean case, and $(1, t)$ for the linear trend case, and $\hat{\psi}$ is the GLS estimator obtained by regressing $\bar{y}$ on $\bar{z}$ where

$$\begin{aligned} \bar{y} &= (y_1, (1 - \bar{\alpha}B)y_2, \ldots, (1 - \bar{\alpha}B)y_T)' \\ \bar{z} &= (z_1, (1 - \bar{\alpha}B)z_2, \ldots, (1 - \bar{\alpha}B)z_T)' \end{aligned}$$

and $\bar{\alpha} = 1 + \bar{c}/T$. They also consider a point optimal test of the unit root null hypothesis $\alpha = 1$ against the alternative $\alpha = \bar{\alpha}$ given by

$$PT = [S(\bar{\alpha}) - \bar{\alpha}S(1)] / s_{ar}^2$$

where $S(a)$ is given by $(y_a - z_a\psi)'(y_a - z_a\psi)$, and s_{ar} is the autoregressive spectral density estimator of the long-term variance. The value of $\bar{c}$ is chosen such that the asymptotic power of test is 50% against the local alternative $(\bar{\alpha} = 1 + \bar{c}/T)$. ERS advise $\bar{c} = -7$ for the constant mean case and $\bar{c} = -13.5$ for the linear trend case.

2.2 Ng and Perron (2001)

Ng and Perron (2001) (NP, thereafter) proposed modifications of the Phillips and Perron (1988) test, which is a non-parametric approach to correct residual autocorrelation by modifying the Dickey-Fuller test statistics, first, to correct the size distortions (as suggested by Perron and Ng, 1996), second, to improve the power (as suggested by ERS, 1996). The NP test is based on the following regression

$$\Delta \tilde{y}_t = \beta_0 \tilde{y}_{t-1} + \sum_{j=1}^{k} \hat{\phi}_j \Delta \tilde{y}_{t-j} + \hat{\varepsilon}_t$$

where $\tilde{y}_t$ is the locally detrended series y_t. Under the unit root null hypothesis, $\beta_0 = 0$; thus the NP test statistics, called M-GLS tests, are

$$
\begin{aligned}
MZ_t &= \left(T^{-1}\tilde{y}_T^2 - s_{ar}^2\right)\left(4s_{ar}^2 T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2\right)^{-1/2} \\
MZ_a &= \left(T^{-1}\tilde{y}_T^2 - s_{ar}^2\right)\left(2T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2\right)^{-1}
\end{aligned}
$$

where s_{ar} is the autoregressive spectral density estimator of the long-term variance. NP also consider a modified feasible point optimal test

$$
MPT = \left[\bar{c}^2 T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2 - \bar{c}T^{-1}\tilde{y}_T\right]/s_{ar}^2
$$

2.3 Kwiatkowski, Phillips, Schmidt and Shin (1992)

We also consider the stationarity test of Kwiatkowski et al. (1992) (KPSS, thereafter). This method tests the null hypothesis of stationarity around a trend against the unit root alternative. The test proceeds as follows: To implement the test, we first obtain the residual $\hat{e}_t$ from the regression of X_t on a constant and a trend

$$
X_t = \alpha + \beta t + e_t
$$

The KPSS test statistic is given by

$$
KPSS = \frac{1}{s_{wa}^2}\frac{\sum_{t=1}^{T}\hat{S}_t^2}{T^2}
$$

where $\hat{S}_t$ is the partial sum process defined by: $\hat{S}_t = \sum_{i=1}^{t}\hat{e}_i$ for $t = (1, ..., T)$, and s_{wa}^2 is the estimator of the long-term variance of $\hat{e}_t$ given by: $s_{wa}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{T-1} w(s,l)\hat{\gamma}_j$ where $\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-j}$, and $w(s,l)$ being an optimal weighting function corresponding to the choice of a spectral window.[8]

3 Business-Cycle Models as Data Generating Processes

We take the models of Chang et al. (2007) as DGPs. This section provides a brief description of the model and of the various specifications studied. The model is real and perfectly competitive. Households

[8]We use a Bartlett window, $w(l, s) = 1 - s/(l + 1)$, as suggested by KPSS (1992).

consume, accumulate physical capital, and supply production factors (labor and physical capital) to firms. The objective of the representative household is to maximize the expected intertemporal utility function

$$\mathrm{E}_t \left\{ \sum_{s=0}^{\infty} \beta^{t+s} \left(\ln C_{t+s} - \frac{\left(H_{t+s}/B_{t+s}\right)^{1+1/\nu}}{1+1/\nu} \right) \right\} \tag{1}$$

where $0 < \beta < 1$ is the subjective discount factor, ν the Frisch elasticity of labor supply, C_t the household consumption, H_t the household hours worked, B_t a preference shock on the desutility of labor, and t the period. The representative household faces the budget constraint

$$W_t H_t + R_t K_t = C_t + K_{t+1} - (1-\delta) K_t \tag{2}$$

where δ is the depreciation rate of physical capital, W_t the wage rate, R_t the rental rate of physical capital, and K_t the stock of physical capital held by the household. The representative firm combines physical capital and labor to produce the final good according to

$$Y_t = \left(A_t H_t^d\right)^{\alpha} \left(K_t^d\right)^{1-\alpha} \left[1 - \varphi \left(\frac{H_t^d}{H_{t-1}^d} - 1 \right)^2 \right] \tag{3}$$

where $0 < \alpha < 1$ is the elasticity parameter of the production function, A_t is the technological shock common to all firms, H_t^d and K_t^d the demand of inputs, and $\varphi \geq 0$ measures the size of the adjustment costs of labor. The model description is closed with the shock processes

$$\ln A_t = \gamma + \ln A_{t-1} + \varepsilon_{a,t},\ \varepsilon_{a,t} \sim iid\left(0, \sigma_a\right) \tag{4}$$

$$\ln B_t = \rho_b \ln B_{t-1} + (1-\rho_b) \ln B_0 + \varepsilon_{b,t},\ \varepsilon_{b,t} \sim iid\left(0, \sigma_b\right) \tag{5}$$

where $\gamma > 0$ is the deterministic component of the technological shocks drift and $0 < \rho_b \leq 1$ is the persistence parameter of the preference shocks.

The model is calibrated using the outcome of the estimations of Chang et al. (Table 2, p. 1366, 2007) for the four specifications given in Table 2. The model is then solved and simulated using the programs provided by the authors.[9]

[9] The required programs are dsge.g, dsgemod.src, and dsgesim.src.

4 Results

To assess the performances of the unit root and stationarity tests with the DGPs described in Table 2, a Monte Carlo study is performed. Table 3 displays the results for the DGPs where hours worked are stationary whereas Table 4 reports the results for the DGPs where the hoiurs worked are non-stationary, without (Panel A) and with (Panel B) adjustment costs. The power of unit root tests and the size of stationarity tests are given in Table 3 whereas the size of unit root tests and the power of stationarity tests are presented in Table 4.[10] The sample sizes considered are $T = 100$, 200, 500 and 1000, and all experiments are based on 30,000 replications. Observed unit-root and stationary test statistics are compared to their finite-sample 5% critical values given in (*i*) the original papers of the unit root tests, (*ii*) MacKinnon (1991), Sephton (1995) and Vougas (2007) for the small finite-sample, and (*iii*) our computations.

For the DGPs with non-stationary hours worked (Table 4), the unit root tests display good size property, whatever the sample sizes, and without and with adjustment costs. Note that the KPSS tests display less power, especially with the model with adjustment costs (Panel B).

For the DGPs with stationary hours worked (Table 3), the major issue is for stationary hours worked on small samples ($T = 100$ and 200), i.e. sample sizes of typical macroeconomic series. In this case, significant differences appear between tests and interestingly also between model specifications. Overall, the NP tests (MZ$_\alpha$, MZ$_t$ and MPT) exhibit higher power than the others studied tests, but with some difference according to the model specification. For the model without adjustment costs (Panel A), the NP tests reject the unit root hypothesis at a rate of 45% (especially for MZ$_\alpha$ and MZ$_t$) against 37% for the ADF test ($T = 200$) and 40% for ERS tests (DF-GLS and PT). This slight difference does not allow to have some indications about the preference of the NP tests rather than the standard unit root test. The conclusion is different for the model with adjustment costs (Panel B). In this case, the NP tests reject the unit root hypothesis at a rate of 72% and 50% against 1% for the ADF test for $T = 200$ and $T = 100$, respectively. Note that the ERS tests have slightly less powerful than the NP tests. Moreover, the KPSS tests display strong size distortions in small samples, especially for $T = 100$. Therefore, it seems that the efficient unit

[10] For the KPSS test, we choice $l = INT[12(T/100)^{1/4}]$. We obtained the same results with $l = INT[8(T/100)^{1/4}]$.

root tests, especially the NP tests, are more powerful than the standard unit root test. This indicates that the NP tests should be preferred to the ADF test in this framework given the fact that the model with adjustment costs is more consistent with empirical facts than the model without adjustment costs, as shown by Chang et al. (2007).

The sensitivity of the test performances to the specification of the model proceeds from the amplification and propagation mechanisms of the chosen model. Adjustment costs are well known to propagate the effects of shocks in the economy. Agents smooth the adjustment of labor to diminish total costs. Since adjustment costs increase the persistence of shocks in the economy, it is surprising that the NP tests reject more frequently the unit root hypothesis than without adjustment costs. To understand this point, it is worthily to make the distinction between the *endogenous persistence*, associated with adjustment costs, and the *exogenous persistence*, associated with the persistence of the exogenous shocks on labor supply. Chang et al. (2007) put forward an inverse relation between the two forms of persistence in their estimation procedure: A high value for φ, which measures the size of adjustment costs, is associated with a low value of ρ_b, which measures the autocorrelation of the labor supply shocks (see Table 2). Figure 1 shows the sharp contrast in responses of hours to both shocks. The model without adjustment costs generates monotonic responses of labor to a stationary supply shock, but which last for a very long time, whereas the model generates hump-shaped responses of labor with a quicker return to the steady state. Hump-shaped behavior in macroeconomic time series is a key features of business cycle (see, e.g., Cogley and Nason, 1995). Besides, Chang et al. (2007) conclude that the model with adjustment costs and stationary hours has the best fit among the four specifications. Given these findings, our results suggest to use the efficient unit root tests proposed by Ng and Perron (2001) because they have higher power than the ADF test when simulated series are hump-shaped.

5 Concluding remark

Mixed results of unit root tests in small samples cast doubts on their usefulness in developing business cycle theory. In this article, we attempted to improve the contribution of unit root tests to economic theory by linking the assessment process of test quality together with economic theory. From Monte Carlo simulations based on a well-specified business cycle model as DGPs, namely the Chang et al. (2007) model with labor adjustment costs, we showed that the efficient unit root tests proposed by Ng and Perron (2001) have higher power than the standard ADF unit root test. This finding suggests that these efficient tests should be preferred in this framework. We think that it would be useful for the macroeconomists that the further developments on unit root tests include business-cycle models as a data generating process to evaluate their properties.

References

Blanchard, O.J., Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. American Economic Review, 79, 655-73.

Chang, Y., Doh, T., Schorfheide, F. (2007). Non-stationary Hours in a DSGE Model. Journal of Money, Credit and Banking, 39, 1357-1373.

Chari, V.V., Kehoe, P.J., McGrattan, E.R. (2008). Are structural VARs with long-run restrictions useful in developing business cycle theory. Journal of Monetary Economics, 55, 1337-52.

Christiano, L.J., Eichenbaum, M. (1992). Current real business cycle theories and aggregate labor market fluctuations. American Economic Review, 82, 430-450.

Christiano, L.J., Eichenbaum, M., Vigfusson, R. (2004). What happens after a technology shock. Working Paper No 9819, NBER.

Cogley, T., Nason, J. (1995). Output dynamics in Real-Business-Cycle models. American Economic Review, 84, 492-511.

Dickey, D.A., Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with unit root. Journal of the American Statistical Association, 74, 427-481.

Dickey, D.A., Fuller, W.A. (1981). "Likelihood ratio statistics for autoregressive time series with unit root. Econometrica, 49, 1057-1072.

Elliott, G., Rothenberg, T.J., Stock, J.H. (1996). Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.

Fève, P., Guay, A. (2009). The response of hours to a technology shock: A two-step structural VAR approach. Journal of Money, Credit and Banking, 41, 987-1013.

Francis, N., Ramey., V.A. (2005). Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited. Journal of Monetary Economics, 52, 1379-99.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? American Economic Review, 89, 249-71.

Galí, J., Rabanal, P. (2004). Technology shocks and aggregate fFluctuations: How well does the real business cycle model fit postwar U.S. data? Working Paper No 10636, NBER.

Haldrup, N., Jansson, M. (2006). Improving size and power in unit root testing. In Mills, T.C., Patterson, K. (eds.), Palgrave Handbook of Econometrics: Econometric Theory, Palgrave Macmillan.

Hansen, G.D. (1985). Indivisible labor and the business cycle. Journal of Monetary Economics, 16, 309-327.

Hansen, G.D. (1997). Technological progress and aggregate fluctuations. Journal of Political Economy, 16, 1005-1023.

King, R., Plosser, C., Stock, J., Watson, M., (1991). Stochastic trends and economic fluctuations. American Economic Review, 81, 819-940.

Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationary against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics, 54, 159-178.

Kydland, F., Prescott, E. (1982). Time to build and aggregate fluctuations. Econometrica, 50, 1345-70.

MacKinnon, J. (1991). Critical values for cointegration tests. In: Engle R. and Granger C. (eds), Long-run Economic Relationships, Readings in Cointegration, Oxford University Press, pp 267–276.

Nelson, C.R., Plosser, C.I. (1982). Trends and random walks in macroeconomic time series. Journal of Monetary Economics 10, 139-162.

Ng, S., Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-1554.

Perron, P., Ng, S. (1996). Useful modifications to unit root tests with dependent errors and their local asymptotic properties. Review of Economic Studies 63, 435-465.

Phillips, P.C.B., Perron, P. (1988). Testing for unit root in time series regression. Biometrika 75, 347-353.

Prescott, E.C. (1986). Theory ahead of business cycle measurement. Carnegie Rochester Conference Series on Public Policy, 25, 11-44.

Sephton, P.S. (1995). Response surface estimates of the KPSS stationarity test. Economics Letters, 47, 255-261.

Vougas, D.V. (2007). GLS detrending and unit root testing. Economics Letters, 97, 222–229.

Whelan, K.T. (2009). Technology shocks and hours worked: Checking for robust conclusions. Journal of Macroeconomics, 31, 231-239.

Tables and Figure

Table 1: Results of unit root tests.

Dataset	MZ_α	MZ_t	DF-GLS	PT	MPT	k^a	ADF^b	k^b
Dataset 1	-12.40*	-2.47*	-2.48*	2.09*	2.05*	1	-2.80	4
Dataset 2	-3.65	-1.34	-1.42	7.93	6.71	1	-2.55	4
Dataset 3	-11.20*	-2.34*	-2.43*	2.29*	2.30*	1	-2.44	4
Critical value at the 5% level	-8.10	-1.98	-1.98	3.17	3.17		-2.86	

Notes: * indicates rejection of the unit-root null hypothesis at the 5% level of significance. [a] the lag order k in the regression is selected by using the Modified Information Criteria (MIC) proposed by Ng and Perron (2001). [b] the values of the ADF tests and lag order k are taken in Chang et al. (footnote 7, p. 1363, 2007). The three datasets have been collected by Chang et al. (2007). Dataset 1 is constructed by the Bureau of Labor Statistics and corresponds to the average weekly hours of all people in the non-farm business sector. Dataset 2 has been constructed by Christiano et al. (2004) (LBMN, DRI-Global Insight). Dataset 3 has been constructed by Gali and Rabanal (2004) and corresponds to non-farm business sector hours (LXNFH, Haver Analytics' USECON). MZ_α, MZ_t and MPT denote the Ng and Perron (2001) tests; DF-GLS and PT denote the Elliot et al (1996) tests; and ADF denotes the Augmented Dickey and Fuller (1981) test.

Table 2: Specifications of the DGPs.

Specification	Hours worked	Adjustment costs	Parameter values
1	Stationary	Without	$\{\rho_b = 0.951; \varphi = 0.000\}$
2	Non-Stationary	Without	$\{\rho_b = 1; \varphi = 0.000\}$
3	Stationary	With	$\{\rho_b = 0.800; \varphi = 11.36\}$
4	Non-Stationary	With	$\{\rho_b = 1.000; \varphi = 8.054\}$

Source: Chang et al. (2007). See Table 2 of Chang et al. (2007) for the other parameters

Table 3: Reject rates of unit-root and stationary test statistics – DGP: stationary hours worked.

Sample	MZ_α	MZ_t	DF-GLS	PT	MPT	ADF	KPSS
Panel A: Without adjustment costs (1)							
$T = 1000$	0.9326	0.9281	0.9341	0.9242	0.9316	1.0000	0.0123
$T = 500$	0.7984	0.8014	0.8320	0.7762	0.7953	0.9601	0.0656
$T = 200$	0.4503	0.4580	0.4089	0.3952	0.4299	0.3688	0.2592
$T = 100$	0.1859	0.1605	0.1460	0.1432	0.1643	0.1326	0.4439
Panel B: With adjustment costs (3)							
$T = 1000$	0.9196	0.9154	0.9314	0.9125	0.9202	0.9997	0.0128
$T = 500$	0.8579	0.8620	0.8659	0.8422	0.8577	0.4216	0.0572
$T = 200$	0.7286	0.7326	0.6821	0.6776	0.7148	0.0125	0.2387
$T = 100$	0.5168	0.4874	0.3977	0.4369	0.4894	0.0083	0.4138

Notes: (1) and (3) denote the specifications 1 and 3 in Table 2. MZ_α, MZ_t and MPT denote the Ng and Perron (2001) tests; DF-GLS and PT denote the Elliot et al (1996) tests; ADF denotes the Augmented Dickey and Fuller (1981) test; and KPSS denotes the Kwiatkowski et al. (1992) test.

Table 4: Reject rates of unit-root and stationary test statistics – DGP: non-stationary hours worked.

Sample	MZ_α	MZ_t	DF-GLS	PT	MPT	ADF	KPSS
Panel A: Without adjustment costs (2)							
$T = 1000$	0.0656	0.0615	0.0683	0.0639	0.0649	0.0620	0.4795
$T = 500$	0.0642	0.0649	0.0658	0.0598	0.0624	0.0568	0.4480
$T = 200$	0.0599	0.0603	0.0491	0.0486	0.0535	0.0590	0.3489
$T = 100$	0.0550	0.0455	0.0403	0.0401	0.0466	0.0549	0.2819
Panel B: With adjustment costs (4)							
$T = 1000$	0.0541	0.0501	0.0536	0.0530	0.0534	0.0390	0.1844
$T = 500$	0.0547	0.0559	0.0535	0.0515	0.0538	0.0391	0.1763
$T = 200$	0.0615	0.0616	0.0424	0.0500	0.0550	0.0389	0.1321
$T = 100$	0.0722	0.0614	0.0390	0.0528	0.0609	0.0446	0.1184

Notes: (2) and (4) denote the specifications 2 and 4 in Table 2. MZ_α, MZ_t and MPT denote the Ng and Perron (2001) tests; DF-GLS and PT denote the Elliot et al (1996) tests; ADF denotes the Augmented Dickey and Fuller (1981) test; and KPSS denotes the Kwiatkowski et al. (1992) test.

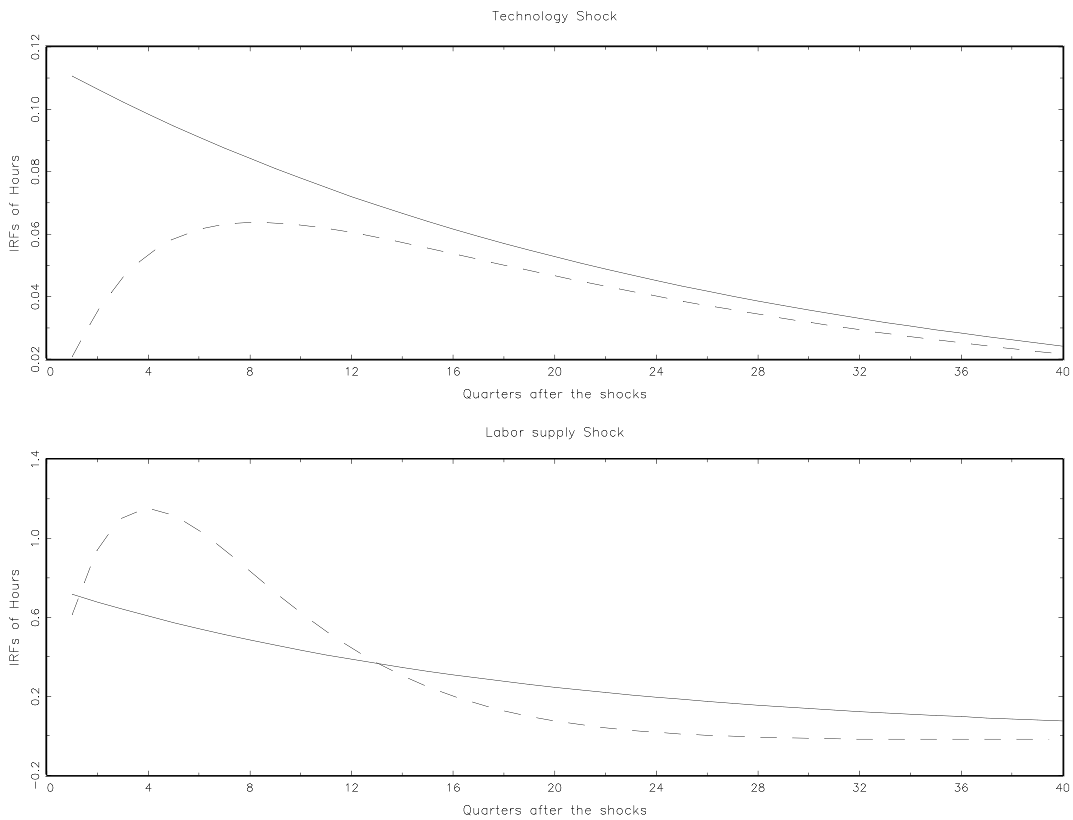

Figure 1: Impulse Response Functions of Hours Worked to shocks with (dotted lines) and without labor adjustment costs (solid lines).